Evidence-based Practice Center Systematic Review Protocol

Project Title: Imaging Tests for the Staging of Colorectal Cancer

I. Background and Objectives for the Systematic Review

Colorectal Cancer

In the United States, each year about 100,000 patients receive a diagnosis of colon cancer and another 50,000 receive a diagnosis of rectal cancer.[1] Colorectal cancer most commonly affects older adults, with 90 percent of cases diagnosed in individuals older than 50 years.[2] Colorectal cancer is often fatal, with approximately 50,000 deaths attributed to it each year in the United States.[1] As such, it is the third most common type of cancer and also the third most common cause of cancer-related death for both men and women. Colorectal cancer is also associated with high health care costs. It has been estimated to be the cancer site with the second-highest associated cost of care (second only to female breast cancer).[3,4]

Although often mentioned together as if they were the same condition, colon and rectal cancer differ significantly in their epidemiology, prognosis, and treatment. Colon cancer is more common than rectal cancer and can be subdivided into proximal (involving the cecum, the ascending colon, and the transverse colon) and distal (involving the descending colon). Men are more likely to develop distal colon and rectal cancer, and women and younger patients of either gender are more likely to develop proximal colon cancer.[5,6]

Colorectal cancer may be diagnosed during screening of asymptomatic individuals or after a patient develops symptoms. Colon cancer symptoms include abdominal discomfort, change in bowel habits, anemia, and weight loss. Rectal cancer symptoms include bleeding, diarrhea, and pain. The U.S. Preventive Services Task Force currently recommends screening for colorectal cancer using fecal occult blood testing, sigmoidoscopy, or colonoscopy, beginning at age 50 years and continuing until age 75 years.[7] Diagnosis is usually established through histopathological examination of tissue samples (obtained through fiber-optic colonoscopy or biopsy).

Staging Systems

Once the diagnosis has been established, patients with colorectal cancer undergo testing to establish the extent of disease spread, which is called staging. Staging is used primarily to determine appropriate treatment strategies; treatment options for colorectal cancer vary widely depending on the stage of disease at diagnosis. Stage is not the only determinant of treatment options—patient comorbidities and treatment preferences and clinician and institution preferences are also used in decisionmaking. However, stage is the key determinant of the management strategy. Staging is also used to inform patient prognosis and to identify patients at higher risk of relapse or cancer-related mortality.

For colorectal cancer there exists a widely accepted "TNM" staging system endorsed by the American Joint Committee on Cancer (AJCC). This system is consistent with the Union for International Cancer Control (UICC) staging system and, therefore, allows for direct comparisons across clinical research centers and countries. The AJCC system aims to characterize the anatomic extent of colorectal cancer based on three tumor characteristics: the extent of tumor infiltration into the bowel wall (tumor stage, designated as "T"), the extent of local or regional lymph node spread (nodal stage, designated as "N"), and the presence of distant metastatic lesions (metastatic spread, designated as "M"). Clinicians determine the T, N, and M components and use them to assign patients into four broad disease stages of increasingly unfavorable prognosis (denoted I through IV). The categories are mutually exclusive (i.e., a patient can belong to only one category) and exhaustive (i.e., all patients belong to a category). There are two other colorectal cancer staging systems—the Dukes and modified Astler-Coller staging systems—that are less widely used. One of the challenges of this systematic review is to determine how cancer stages can be translated between staging systems or within versions of the AJCC staging system, currently in its seventh edition. The three staging systems are summarized in Appendix A.

Staging and Interim Restaging

Staging can be performed at two distinct time points in the management of colorectal cancer. The first is immediately after diagnosis, before any treatment has been given. Imaging, clinical examination, and biomarker assessment are used to assign a stage, which is used to make decisions about primary treatment and management. The second time point applies only to patients who, on the basis of their primary staging, were treated with neoadjuvant chemotherapy or radiotherapy instead of by immediate surgery. Staging after treatment (interim staging, or restaging) is primarily intended to determine if the tumor has responded to the treatment (downstaging). The role of imaging at each of these two time points is very different, and we will address them in separate Key Questions (KQs) in this review.

For stage I, II, and III disease, surgical resection is the primary treatment. Patients with stage III colon cancer are usually also treated with adjuvant chemotherapy; controversy exists over whether stage II patients should also receive adjuvant chemotherapy. For patients with stage II or III rectal cancer, preoperative chemotherapy and possibly radiation is the preferred treatment. Surgery is an option for some patients with stage IV colorectal cancer, but for these patients, primary treatment is chemotherapy.[8] Patients who receive systemic treatments before undergoing surgical resection may have their disease restaged after their treatment (interim staging) to adjust treatment, accounting for any effect of the systemic treatment on the tumor.

Recurrent Colorectal Cancer

Recurrent colorectal cancer arises in some patients after they have undergone apparently successful initial treatment for primary colorectal cancer. Approximately 20% to 30% of patients develop recurrent disease. After completion of primary treatment, patients usually enter a routine surveillance program intended to detect signs of recurrence. Typically this consists of regular tests for biomarkers (such as carcinoembryonic antigen), clinical examination, colonoscopies, and, possibly, computed tomography (CT) scans.[8]

After the diagnosis of a recurrence, staging aims to assess the extent of disease to guide treatment decisions and determine prognosis. Multiple treatment options (e.g., chemotherapy alone vs. multimodality therapy including metastasectomy) are available for patients with recurrent disease, and the choice between them is primarily based on accurate assessment of the extent of disease.[8]

Imaging for Colorectal Cancer Staging

Staging tests can be crudely divided into noninvasive and invasive tests. Invasive tests include surgical sampling of lymph nodes for histopathological evaluation and intraoperative or laparoscopic ultrasound. Although certain aspects of staging (such as histopathological examination of sampled lymph nodes) require invasive tests, staging tests using noninvasive means are also important. For example, decisions about which patients should receive presurgical chemotherapy treatment require input from noninvasive imaging. This review will focus on noninvasive imaging tests for staging.

Imaging tests can be broadly divided into two categories—some tests primarily provide anatomical information (e.g., CT), while others primarily provide functional information in terms of metabolic activity (e.g., positron emission tomography [PET]). Table 1 summarizes the imaging modalities that are currently available for colorectal cancer staging. An important characteristic of imaging tests is whether they use ionizing radiation; for patients with colorectal cancer who have a long life expectancy (e.g., those with early stage disease who undergo treatment with curative intent), the cumulative exposure to ionizing radiation during diagnosis, staging, and subsequent surveillance can be substantial.[9]

Table 1. Imaging tests that have been used in colorectal cancer staging

Imaging Modality (Abbreviation)	Key Characteristics of the Imaging Technology	Use of Ionizing Radiation
Chest x-ray	A chest X-ray is an anatomical imaging technique that can be used to detect pulmonary metastatic disease. It has substantially lower resolution than CT or MRI and, thus, has lower sensitivity. It will not be evaluated in this review.	Yes
Computed tomography (CT)	A cross-sectional imaging technique that uses x-rays to provide detailed anatomical information of any region of the body. CT can be performed with or without an intravenous contrast agent.	Yes
Multidetector computed tomography (MD-CT)	MD-CT scanners are based on the same imaging principles as conventional CT devices, but they acquire multiple sections simultaneously (by having a 2-dimensional array of detectors), thus increasing the speed of image acquisition. MD-CT scanners provide high-resolution anatomical information.	Yes
CT colonography	Sometimes called “virtual colonoscopy,” this anatomical imaging method uses a CT scanner to create a 3-dimensional image of the interior of the colon (as well as the surrounding tissues). The colon is usually inflated with air for this procedure.	Yes

Imaging Modality (Abbreviation)	Key Characteristics of the Imaging Technology	Use of Ionizing Radiation
Magnetic resonance imaging (MRI)	An imaging technique that uses a magnetic field to obtain anatomical images of the body. MRI scanning is slower than CT scanning and requires the patient to remain still during image acquisition. It is particularly useful for imaging the rectum (because the rectum is held relatively fixed by surrounding tissues) or liver. MRI offers primarily anatomical information, but specific techniques (e.g., diffusion-weighted imaging) can also provide functional information. Additionally, MRI machine attachments allow visualization of certain anatomical structures (e.g., an endorectal coil for rectal cancer).	No
Transabdominal ultrasound (TUS)	This imaging test uses ultrasound waves and their reflection from internal organs to provide images of the underlying anatomy. In the setting of colorectal cancer staging, TUS is mainly used to image the liver for the potential presence of metastases.	No
Endoscopic ultrasound (EUS) and transrectal ultrasound (TRUS)	These are imaging tests that require the placement of the ultrasound probe within the bowel. They allow imaging of the primary tumor (including the depth of invasion into the bowel wall) and its association with the surrounding tissues.	No
Positron emission tomography (PET)	This is a functional imaging technique that uses radioisotope-tagged tracers to examine the level and type of biochemical activity in lesions suspected to be cancerous. The most commonly used tracer is ^{18}FDG, which detects cells demonstrating increased glucose transport and metabolism (cancer cells exhibit such metabolic activity). Alternative tracers have been investigated for various types of cancer, but ^{18}FDG remains the one most widely available. Although PET can identify cancer lesions before they can be detected by anatomical imaging methods, PET images have lower resolution than CT or MRI. Further, PET exams can produce false-negative results in metabolically inactive tumors or false-positive results in the presence of inflammation. Functional data from PET with anatomical data from CT colonography can be used to improve the accuracy of tumor staging.	Yes
Positron emission tomography combined with computed tomography (PET/CT fusion) or magnetic resonance imaging (PET/MRI fusion)	These methods combine the anatomical information of CT with the functional information of PET by superimposing the two scans. PET/CT fusion overcomes the limitations of PET-only scans, which have relatively low resolution. Some studies use MRI instead of CT as the anatomical technique.	Yes*

* For PET/CT exposure to ionizing radiation is necessary for both test components (PET and CT); for PET/MRI exposure to ionizing radiation is only from the PET scan.

Abbreviation: ^{18}FDG = fluorodeoxyglucose (18F)

The tests of interest can affect the staging evaluation of patients in different ways (i.e., not all tests affect all components of the TNM classification), depending on their technical characteristics. For example, endoscopic ultrasound can provide information on the "local stage" (i.e., the depth to which the cancer invades the bowel wall) but not on the presence of distant metastases. In contrast, whole-body CT or PET/CT can provide information on metastatic lesions, even when they are asymptomatic. Further, no single test may be sufficient for staging, and different combinations of tests may be needed.

Objectives of This Systematic Review

We have summarized key recommendations from organizations in the United States regarding the use of imaging tests for staging in Table 2. As can be seen from the table, the organizations are not in complete agreement about which modalities should be emphasized for the clinical situations described. Also, there is no consensus guidance about the sequence in which these tests are to be applied in the staging process.

The imaging modalities vary in their accuracy and in the harms they can potentially cause. To be clinically useful and relevant, these benefits should be weighed against the potential harms of using the modality. The size of the tumor may also have a significant effect on the accuracy of the imaging modality. For example, American College of Radiology guidelines provide separate recommendations for large and small rectal cancer lesions but National Comprehensive Cancer Network guidelines do not make that distinction. The differences in the testing protocols associated with different imaging modalities can affect their test performance and need to be systematically reviewed. Although it is necessary to identify the most accurate test (or combination of tests) for correctly establishing the stage of the cancer, it is also important to assess the relative impact of testing strategies using different imaging modalities on intermediate outcomes such as stage reclassification (i.e., an indication of how much additional information is obtained by applying a test) and therapeutic decisionmaking (i.e., measures of the impact of tests on clinical decisions) and clinical outcomes. Building our work on the available scientific data, we hope that this systematic review of the available imaging modalities for colorectal cancer staging will uncover evidence to support these questions or highlight issues not addressed by the currently available evidence that may represent targets for future research.

Accurate staging of colorectal cancer is important for determining prognosis and selecting the most appropriate treatment. Selection of more-appropriate treatment options would be expected to improve clinical outcomes (e.g., by avoiding unnecessarily aggressive treatments for low-risk disease). Besides assisting in treatment selection, staging also provides important prognostic information about chances of short- and long-term survival. Our objective in this review is to synthesize the available information on using imaging for staging. This information will help clinicians select protocols for staging, may reduce variability across treatment centers in staging protocols, and may improve patient outcomes.

Table 2. Summary of existing guidelines for staging colorectal cancer

Clinical Description	ACR Recommendations	NCCN Recommendations
Colon cancer	*Usually appropriate* • CT of the chest-abdomen-pelvis with or without contrast • X-ray of the chest (if chest CT is not performed) • FDG-PET of the whole body • MRI of the abdomen and pelvis with or without contrast	*Recommended* • Chest/abdominal/pelvic CT with IV and oral contrast
	May be appropriate • MRI of the abdomen and pelvis without contrast • CT of the chest-abdomen-pelvis with and without contrast • CT of the chest-abdomen-pelvis without contrast	
	Usually not appropriate • None reported	*Usually not indicated* • PET scan • PET-CT does not supplant a contrast-enhanced diagnostic CT

Clinical Description	ACR Recommendations	NCCN Recommendations
Rectal cancer	*Usually appropriate for small lesions* • Endorectal US of the pelvis • X-ray of the chest (if chest is not imaged by CT) • CT of the chest-abdomen-pelvis with or without contrast • MRI of the pelvis with or without contrast *Usually appropriate for large lesions* • X-ray of the chest • CT of the chest-abdomen-pelvis with or without contrast • MRI of the abdomen with or without contrast • MRI of the pelvis with or without contrast • FDG-PET of the whole body	*Recommended* • Chest/abdominal/pelvic CT • Endorectal US or endorectal/pelvic MRI
	May be appropriate for small lesions • FDG-PET of the whole body • MRI of the abdomen with and without contrast • MRI of the abdomen without contrast • CT of the chest-abdomen-pelvis without contrast • CT of the chest-abdomen-pelvis with and without contrast • MRI of the pelvis without contrast *May be appropriate for large lesions* • Endorectal US of the pelvis • MRI of the abdomen without contrast • MRI of the abdomen with contrast • CT of the chest-abdomen-pelvis without contrast • CT of the chest-abdomen-pelvis with and without contrast	
	Usually not appropriate • None reported	*Usually not indicated* • PET-CT not routinely indicated
Suspected liver metastases after detection of primary tumor[11]	*Usually appropriate* • CT of the abdomen with contrast • MRI of the abdomen with and without contrast • FDG-PET from the skull base to mid-thigh *May be appropriate* • MRI of the abdomen without contrast • CT of the abdomen with and without contrast • CT of the abdomen without contrast • US of the abdomen *Usually not appropriate* • CTA of the abdomen with contrast • In-111 somatostatin receptor scintigraphy	
Suspected or proven metastatic synchronous adenocarcinoma (M1)		*Recommended* • Chest/abdominal/pelvic CT (with IV contrast) • Consider MRI with IV contrast if CT is inadequate • Needle biopsy (if indicated) *May be appropriate* • PET-CT scan only if M1 disease is potentially curable

Abbreviations: ACR = American College of Radiology; CT = computed tomography; CTA = CT angiography; FDG-PET = fluorodeoxyglucose (F18)/positron emission tomography; IV = intravenously administered; MRI = magnetic resonance imaging; NCCN = National Comprehensive Cancer Network; PET = positron emission tomography; PET-CT = positron emission tomography combined with computed tomography; US = ultrasonography

II. The Key Questions

The draft KQs were posted for public comment in from November 1, 2012 to November 29, 2012, on the Effective Health Care Program Web site. No comments were received; therefore, no changes were made to the KQs. The KQs are listed below:

Key Question 1

What is the comparative effectiveness of imaging techniques for pretreatment staging of patients with primary and recurrent colorectal cancer?

a. What is the test performance of the imaging techniques used (singly, in combination, or in a specific sequence) to stage colorectal cancer when compared with a reference standard?
b. What is the impact of alternative imaging techniques on intermediate outcomes, including stage reclassification and changes in therapeutic management?
c. What is the impact of alternative imaging techniques on clinical outcomes?
d. What are the adverse effects or harms associated with using imaging techniques, including harms of test-directed management?
e. How is the comparative effectiveness of imaging techniques modified by the following factors:
    i. Patient-level characteristics (e.g., age, sex, body mass index)
    ii. Disease characteristics (e.g., tumor grade)
    iii. Imaging technique or protocol characteristics (e.g., use of different tracers or contrast agents, radiation dose of the imaging modality, slice thickness, timing of contrast)

Key Question 2

What is the comparative effectiveness of imaging techniques for restaging patients with primary and recurrent colorectal cancer after initial treatment?

a. What is the test performance of the imaging techniques used (singly, in combination, or in a specific sequence) to restage colorectal cancer when compared with a reference standard?
b. What is the impact of alternative imaging techniques on intermediate outcomes, including stage reclassification and changes in therapeutic management?
c. What is the impact of alternative imaging techniques on clinical outcomes?
d. What are the adverse effects or harms associated with using imaging techniques, including harms of test-directed management?
e. How is the comparative effectiveness of imaging techniques modified by the following factors:
    i. Patient-level characteristics (e.g., age, sex, body mass index)
    ii. Disease characteristics (e.g., tumor grade)
    iii. Imaging technique or protocol characteristics (e.g., use of different tracers or contrast agents, radiation dose of the imaging modality, slice thickness, timing of contrast)

PICOTS Criteria

Populations

- Adult patients with an established diagnosis of primary colorectal cancer
- Adult patients with an established diagnosis of recurrent colorectal cancer

Interventions

Noninvasive imaging using the following tests (alone or in combination) to assess the stage of colorectal cancer:

- CT
- PET/CT
- MRI
- Endoscopic ultrasound

Combinations of particular interest include endoscopic ultrasound to evaluate the T stage combined with PET/CT or CT to evaluate the N and M stages.

Reference Standards To Assess Test Performance

- Histopathological examination of tissue
- Intraoperative findings
- Clinical followup

Histopathology of surgically resected specimens is the reference standard for pretherapy staging. In patients undergoing surgery, the nodal (N) stage and spread of the tumor to nearby regional structures and other organs is assessed intraoperatively, either by palpation or ultrasound. However, in patients with metastatic disease who undergo palliative care, a combination of initial biopsy results and clinical followup serves as the reference standard.

Clinicians use the results from the imaging modality or modalities to arrive at a stage determination that is compared against the stage established by the reference standard. These comparisons tell us how many people were correctly classified in the various stages of the disease and allow us to calculate the test performance metrics of sensitivity, specificity, and accuracy. The selection of the reference standard is important in evaluating the true performance of an imaging modality for staging.

Comparators

- Any direct comparisons of the imaging tests of interest
- Any direct comparisons of variations of any of the imaging tests of interest (e.g., diffusion-weighted MRI vs. T2-weighted MRI)

Comparators thought to be of particular clinical interest are listed below:

- For colon cancer: a contrast-enhanced CT of the chest, abdomen, and pelvis versus whole-body PET/CT versus a contrast-enhanced MRI of the chest, abdomen, and pelvis
- For rectal cancer: a contrast-enhanced CT of the abdomen and pelvis versus an MRI of the abdomen and pelvis
- For rectal cancer: endoscopic ultrasound versus MRI
- For suspected liver metastasis: CT scan versus MRI or PET/CT of the abdomen
- For suspected widespread metastasis, CT of the chest, abdomen, and pelvis versus whole-body PET/CT or contrast-enhanced MRI of the chest, abdomen, and pelvis

We note that this list is based on a preliminary literature search and discussions with a limited number of clinicians and the Technical Expert Panel (TEP). Thus, we do not anticipate that the listed items cover all of the comparisons of interest. We expect that additional comparisons will be identified during the literature review.

Outcomes

- Test performance outcomes
  - Test performance (e.g., sensitivity, specificity, understaging, and overstaging) against a reference standard test (pathological examination, intraoperative findings, clinical followup)
- Intermediate outcomes
  - Stage reclassification
  - Changes in therapeutic management
- Clinical outcomes
  - Overall mortality
  - Colorectal cancer–specific mortality
  - Quality of life and anxiety
  - Need for additional staging tests, including invasive procedures
  - Need for additional treatment, including surgery, radiotherapy, or chemotherapy
  - Resource utilization related to testing and treatment (when reported in the included studies)
- Adverse effects and harms

- Harms of testing per se (e.g., radiation exposure)
- Harms from test-directed treatments (e.g., overtreatment, undertreatment)

Timing

- Primary staging
- Interim restaging
- Duration of followup will vary by outcome (e.g., from no followup for test performance measurements to many years for mortality)

Setting

- Any setting will be considered.

III. Analytic Framework

An analytical framework illustrating the connections between the populations of interest, the staging modalities, and the outcomes is shown in Figure 1. Note the patient populations of interest are patients newly diagnosed with cancer or newly diagnosed with recurrent cancer. Populations that have completed treatment for colorectal cancer and are undergoing surveillance for recurrences are outside the scope of this report, as are asymptomatic individuals who are undergoing screening. The use of imaging in diagnosing colorectal cancer is also outside the scope of this report.

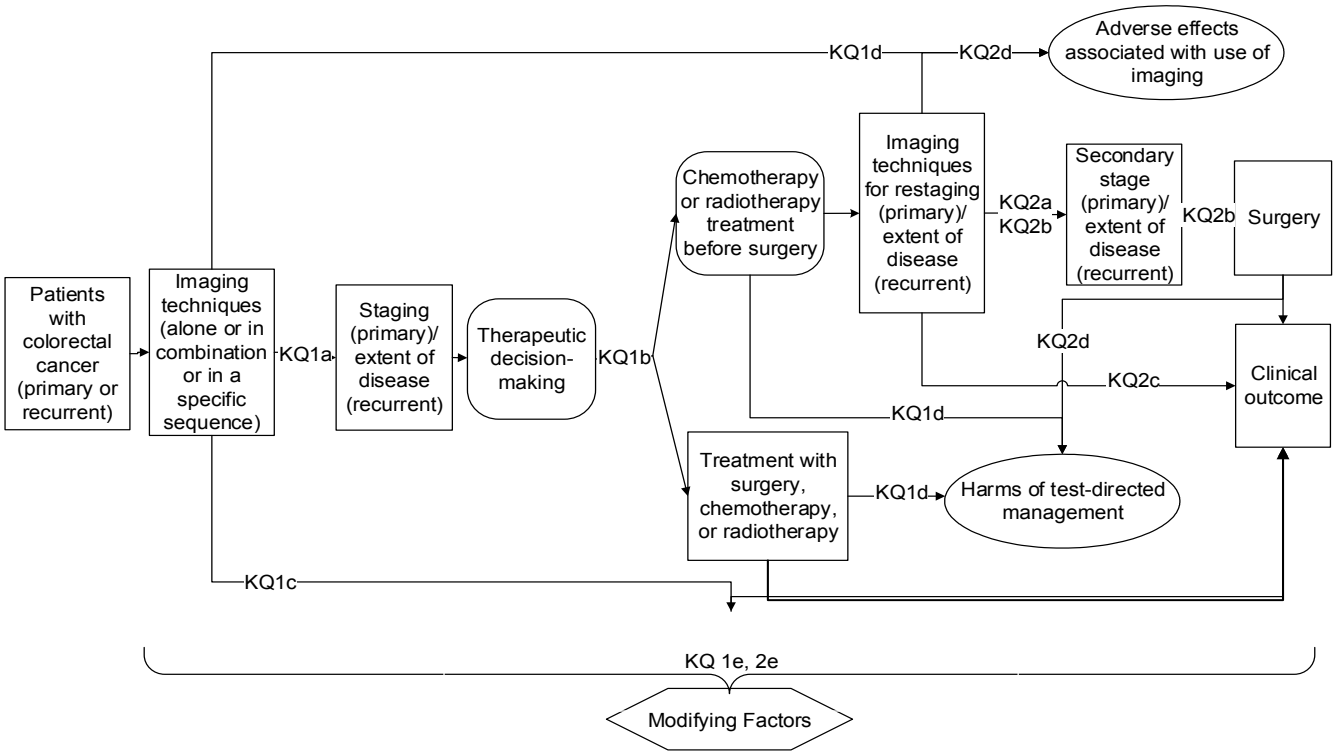
The populations of interest enter the diagram at the left, undergo primary staging (KQ 1), and then commence treatment. Some patients also undergo restaging after completing presurgical treatments such as chemotherapy (KQ 2), and then proceed with the rest of their treatment. Intermediate outcomes such as test performance and harms of testing can be measured immediately after performing the tests, but many of the relevant patient-oriented outcomes (such as mortality) can only be measured after completion of treatment. The point in the process at which each KQ is most relevant is shown on the figure by the placement of the KQ number (1 or 2) and subpart (e.g., a, b, c). The modifying factors affecting test performance in both the primary staging and restaging settings are shown in a separate box at the bottom of the figure.

Although not specified in the figure for simplicity, the four primary patient populations will be considered separately—primary versus recurrent disease and primary staging versus interim restaging. If the data permit, additional groups will also be considered separately—rectal versus colon cancer, proximal colon versus distal colon cancer, and lower rectal versus middle rectal versus upper rectal cancer.

An important factor in selecting an imaging modality for staging is the availability of that modality. Although this factor will not be addressed formally in the review via a KQ, we plan to collect and provide relevant information about the availability and accessibility of imaging modalities, and, if available, information about current patterns of care. This information will be

presented in the background and discussion sections to help place the evidence review findings in context.

Figure 1. Analytical framework of colorectal cancer staging review



KQ = key question

IV. Methods

A. Criteria for Study Inclusion and Exclusion

As suggested in the Agency for Healthcare Research and Quality (AHRQ) *Methods Guide for Effectiveness and Comparative Effectiveness Reviews* (hereafter *Methods Guide*), the inclusion criteria are listed below in separate categories pertaining to publication type, study design, patient characteristics, test characteristics, and reported data.[12]

Publication Criteria

Full-length articles. The article must be published as a full-length, peer-reviewed study. Abstracts and meeting presentations will not be included because they do not include sufficient details about experimental methods to permit an evaluation of study design and conduct; they may also only contain a subset of measured outcomes.[13,14] Additionally, it is not uncommon for abstracts that are published as part of conference proceedings to have inconsistencies when compared with the final study publication or to describe studies that are never published as full articles.[15-19]

Redundancy. To avoid double-counting of patients, when several reports of the same or overlapping groups of patients are available, only outcome data from the report with the largest number of patients will be included. We will make an exception and include data from a smaller study when it reports data on an outcome that was not provided by the largest report or reports longer followup data for an outcome.

English language. Moher et al.[20] have demonstrated that exclusion of non–English language studies from meta-analyses has little impact on the conclusions drawn. Juni et al.[21] found that excluding non–English-language studies typically had little effect on effect size estimates in the majority of meta-analyses they examined, even if those studies were at low risk of bias. Although we recognize that in some situations exclusion of non–English-language studies could lead to bias, we believe that the few instances in which this may occur do not justify the time and cost typically necessary for translating studies.

Study Design Criteria

Single test performance. For questions about the performance of a single imaging test against a reference standard, we will use a three-stage inclusion process. We will first include only recent (2009 or later) high-quality systematic reviews. If there are no recent systematic reviews to support an estimate of test performance for a particular imaging test we will include older high-quality systematic reviews and include only primary studies published after the latest search date in these systematic reviews. If there are no published systematic reviews to support an estimate of test performance for a particular imaging test, we will include primary studies published between 1980 and the present.

Test performance. For questions about test performance, studies of any design—randomized, cross-sectional, case-control, or cohort—will be considered for inclusion. Both retrospective and prospective studies will be considered for inclusion, but retrospective studies must use consecutive enrollment or enrollment of a random sample of participants. Studies must compare the test or test combination being evaluated to a reference standard (see below); for studies comparing two (or more) tests or test combinations, all tests or test combinations must be compared to the same reference standard.

Stage reclassification or clinical decision impact. For questions about stage reclassification or impact on clinician decisionmaking, cross-sectional, cohort, or prospective comparative (randomized or nonrandomized) studies will be considered for inclusion. Studies must directly compare the test or test combination to another test or test combination of interest.

Clinical outcomes. For questions about the impact of testing on patient-oriented clinical outcomes, comparative studies (randomized or nonrandomized) will be considered for inclusion. Studies must directly compare the test or test combination to another test or test combination of interest.

Harms. The adverse events and harms reported by any studies included to address any of the other questions will be used to address questions about harms and adverse events. Additionally, we will search specifically for reports of harms and adverse events associated with each specific imaging modality, such as radiation exposure and reactions to contrast agents. Any study design, including modeling, will be acceptable for inclusion for questions about harms.

Patient Criteria

Type of patient. To be included, the study must have reported data obtained from groups of patients in which at least 85 percent of the patients were from one of the four patient populations of interest. These populations are: (1) patients with newly diagnosed colorectal disease underdoing primary staging; (2) patients with newly diagnosed colorectal disease undergoing interim restaging; (3) patients with newly diagnosed recurrent colorectal disease undergoing primary staging; and (4) patients with newly diagnosed recurrent colorectal disease undergoing interim restaging. Patients must have received a diagnosis of colorectal cancer before enrolling in the study.

Although we have grouped all colon and rectal cancers together as "colorectal cancer" as an inclusion criterion, colon cancer and rectal cancer are somewhat different diseases. Specifically in regard to staging, rectal cancer tends to spread locally and colon cancer tends to spread via distant metastases. Therefore, for accurate staging, colon cancer imaging should focus more on identifying metastases in addition to tumor size and extent; whereas for rectal cancer, imaging of distant metastases is not as important as is gauging tumor depth and local spread. Although we are not requiring that studies report on only rectal or colon cancer for inclusion in the report, when possible (as permitted by the reported data), we will analyze the data for each subgroup—rectal cancer and colon cancer—separately. The location of the rectal tumor—low, middle, or high—may also affect staging accuracy, so when possible we will analyze the data by subgroups of rectal tumor location as well. Some evidence also exists that suggests proximal colon cancer and distal colon cancer may also be distinctly different conditions; so if possible, we will try to analyze data separately by proximal and distal subgroups.[22]

Adults. Only studies of adult patients (older than 17 years of age) will be considered for inclusion.

Test Criteria

Type of test. Only studies of the tests or comparisons of interest (see the Key Questions section for a list) will be considered for inclusion.

Obsolete technology. Experts in imaging technologies constantly innovate, perform research, and devise improvements in technology. Therefore, a need exists to identify and avoid technologies

that have fallen out of routine clinical practice. Using a single cut-off date (e.g., 2001) as a mechanism to eliminate obsolete technology is not thought to be appropriate. Instead, we consulted the TEP about what imaging technologies and variants of imaging technologies are now obsolete and not relevant to clinical practice. The imaging technologies that were determined to be "obsolete" for staging colorectal cancer are the following: transabdominal ultrasound, MRI using endorectal coils, non-multidetector CT, CT arterial portography, CT angiography, CT colonography, and stand-alone PET.

Likewise, experimental technology and prototypes will be excluded. The TEP indicated that PET/MRI and PET/fused with CT colonography are considered experimental at present. MRI using ultrasmall paramagnetic iron oxide is also considered experimental.[23]

Data Criteria

The study must report data pertaining to one of the outcomes of interest (see the Key Questions section for a list).

We will include data from time points and outcomes reported from groups of patients with at least 10 patients with the condition of interest who represent at least 50% of the patients originally enrolled in the study.

B. Searching for the Evidence: Literature Search Strategies for Identifying Relevant Studies To Answer the Key Questions

Literature searches will be performed by Medical Librarians within the Evidence-based Practice Center (EPC) Information Center and will follow established systematic review protocols. We will search the following databases using controlled vocabulary and text words: EMBASE®, MEDLINE®, PubMed®, and The Cochrane Library.

The following gray literature sources will be searched using text words: ClinicalTrials.gov, Centers for Medicare and Medicaid (CMS) Medicare Coverage Database, ECRI Health Devices, Healthcare Standards, the Internet, Medscape, the National Guideline Clearinghouse™ (NGC), and the U.S. Food and Drug Administration (FDA) Web site. We will also use resources available through the EPC Scientific Resource Center (SRC) to access Scientific Information Packets provided by device manufacturers.

An example search strategy is shown in Appendix B.

Literature screening will be performed in duplicate using the database Distiller SR (Evidence Partners, Ottawa, Canada). Literature search results will initially be screened for relevancy. Relevant abstracts will be screened against the inclusion and exclusion criteria in duplicate. Studies that appear to meet the inclusion criteria will be retrieved in full and screened again in duplicate against the inclusion and exclusion criteria. All disagreements will be resolved by consensus discussion among the two original screeners.

The literature searches will be updated during the peer review process before finalizing the review.

C. Data Abstraction and Data Management

Data will be abstracted using the database Distiller SR (Evidence Partners, Ottawa, Canada). Data abstraction forms will be constructed in Distiller and the data will be abstracted into these forms. Duplicate abstraction will be used to ensure accuracy. All discrepancies will be resolved by consensus discussion among the two original abstracters.

Elements to be abstracted include general study characteristics, patient characteristics, details of the imaging methodology, risk of bias items, and outcome data.

D. Assessment of Methodological Risk of Bias of Individual Studies

Studies will be rated as having "Low," "Medium," or "High" risk of bias. The rating will be defined by selecting critical questions from the rating scale that must be answered as "Yes." The critical questions for these ratings for this review were selected after discussions with the TEP.

For studies of test performance, we will use an internal validity rating scale for diagnostic studies to assess the risk of bias of each individual study. This instrument is based on a modification of the QUADAS instrument with reference to empirical studies of design-related bias in diagnostic test studies.[24-26] Each question in the instrument addresses an aspect of study design or conduct that can help to protect against bias, such as enrolling consecutive or a random sampling of patients or blinding image readers to clinical information about the patient. Each question can be answered "Yes," "No," or "Not reported," and each is phrased such that an answer of "Yes" indicates that the study reported a protection against bias on that aspect. The instrument is shown in Appendix C.

For a diagnostic study to be rated as having low risk of bias, the following questions (see the instrument in Appendix C) must all be answered "Yes": questions 1 and 3 (patient enrollment methods), question 6 (blinding of readers), and question 10 (avoided verification bias), and at least six of the other questions must be answered "Yes." The trial will be rated as having high risk of bias if all four of the critical questions are answered "No." The trial will be rated as having moderate risk of bias if it does not meet the criteria for either low or high.

For studies addressing clinical outcomes, we will use an internal-validity rating scale for comparative studies to assess the risk of bias of each individual study. This instrument was developed by the ECRI Institute[27] with reference to empirical studies of the impact of study design on bias in comparative studies and is consistent with the guidance in the AHRQ *Methods Guide*.[28] Each question in the instrument addresses an aspect of study design or conduct that can help to protect against bias, such as randomization of group assignment or blinding outcome assessors to patient group assignment. Each question can be answered "Yes," "No," or "Not Reported," and each question is phrased such that an answer of "Yes" indicates that the study reported a protection against bias on that aspect. The instrument is shown in Appendix C.

For a controlled/comparative study to be rated as having low risk of bias, the following questions (see the ECRI instrument in Appendix C) must all be answered "Yes": questions 1, 2, and 4 (appropriately randomized or used methods to enhance group comparability) and questions 6 and 7 (group comparability), and at least 10 of the other questions must be answered "Yes." The trial will be rated as having high risk of bias if all five of the critical questions are answered

"No." The trial will be rated as Moderate risk of bias if it does not meet the criteria for either Low or High.

As suggest by the AHRQ *Methods Guide*, systematic reviews used to address KQs 1a and 2a will be evaluated for risk of bias with a modified AMSTAR instrument.[29] The instrument is shown in Appendix C.

Systematic reviews will be rated as either "High Quality" or "Not." The rating will be defined by selecting critical questions from the rating scale that must be answered as "Yes." The critical questions for these ratings for this review were selected after discussions with the TEP. For a systematic review to be rated as high in quality, the following questions (see the modified AMSTAR instrument in Appendix C) must all be answered "Yes": questions 2 and 2a (search methods); questions 4 and 4a (study inclusion); questions 7, 7a, and 7b (rating of study quality and strength of evidence); question 8 (methods of analysis); and question 10 (conflicts of interest). Only high-quality systematic reviews will be included to address KQs 1a and 2a.

E. Data Synthesis

For questions addressing individual test performance (accuracy), we will draw evidence from earlier systematic reviews as it is available and only consult the primary literature as is necessary to update or supplement the information available in earlier systematic reviews. As recommended in the AHRQ *Methods Guide*, we will summarize all of the relevant high-quality reviews.[29]

For comparative questions, we will synthesize the evidence from the primary studies themselves. We plan to perform meta-analysis wherever appropriate and possible. Decisions about whether meta-analysis is appropriate will depend on the judged clinical homogeneity of the different study populations, imaging and treatment protocols, and outcomes. When meta-analysis is not possible (due to limitations of reported data) or is judged to be inappropriate, the data will be synthesized using a descriptive, narrative review approach.

For studies of clinical outcomes, we will compute effect sizes and measures of variance using standard methods and will perform DerSimonian and Laird random-effects meta-analysis using Comprehensive Meta-Analysis (CMA) software (Biostat, Inc., Englewood, NJ).

For studies of test performance, we will meta-analyze the data reported by the studies using a bivariate mixed-effects binomial regression model as described by Harbord et al.[30] All such analyses are computed by the STATA 10.0 statistical software package using the "midas" command.[31] In cases in which a bivariate binomial regression model cannot be fit, we will meta-analyze the data using a random-effects model and the software package Meta-Disc.[32]

Meta-regression and subgroup analysis will be used to explore possible causes of heterogeneity. Potential covariates include population descriptors, tumor site and type, country, and setting of care.

Because of the indirect association of staging tests with clinical outcomes and the rarity of studies directly comparing alternative test-and-treat strategies, we anticipate that we may have to piece together evidence on indirect outcomes (test accuracy, impact of tests on stage reclassification, and impact of tests on therapeutic decisionmaking) with direct evidence on clinical outcomes. If we do this, we will describe the indirect chain of evidence in a qualitative fashion.

F. Grading the Strength of Evidence for Individual Comparisons and Outcomes

We will use a formal grading system that conforms with recommendations from AHRQ methods guides (on medical interventions and medical tests) on grading the strength of evidence.[12,33]

We will grade the overall strength of evidence supporting each major conclusion as "High," "Moderate," "Low," or "Insufficient." The grade is developed by considering four important domains: the risk of bias in the evidence base, the consistency of the findings, the precision of the results, and the directness of the evidence. We will assess the risk of bias of each individual study (see the section on assessing the risk of bias of individual studies) to assess its risk of bias for each outcome and will use the aggregate risk of bias to rate the entire evidence base for the comparison and outcome. We will rate the consistency of conclusions by examining individual study-level differences in the direction of effect, differences in point estimates, the degree of overlap of confidence intervals, and, if the conclusions are supported by meta-analyses, with the statistic I^2.[34,35] Data sets found to be inconsistent will be examined for possible causes of the inconsistency; if any are found, subgroup analyses will be performed to generate more than one conclusion, each with its own strength-of-evidence rating. We will use the width of the 95-percent confidence intervals around the summary effect sizes to evaluate the precision of the evidence. Although the outcome of test accuracy is only indirectly related to health outcomes, and because the KQ asks directly about test accuracy, we will grade conclusions about diagnostic test accuracy as being "Direct" for that particular situation.

Publication bias will be addressed by noting the presence of abstracts or ClinicalTrials.gov entries describing studies that did not subsequently appear as full-published articles.

G. Assessing Applicability

The applicability of the evidence involves four key aspects: patients, tests/interventions, comparisons, and settings. After discussions with the TEP, we concluded that the age and sex of patients is unlikely to affect the accuracy of staging, but other patient characteristics—such as race, obesity, genetic syndromes that predispose affected individuals to colorectal cancer, and enrollment of populations with high rates of comorbid conditions—could affect the applicability of study findings, particularly with regard to patient-oriented outcomes. After consulting with the TEP, we addressed tests, interventions, and comparisons by excluding obsolete and experimental imaging tests from inclusion in the report. Settings of care will be described, and if data permit, subgroups of studies will be analyzed separately by setting.

V. References

1. Cancer Facts & Figures 2012 [Internet]. American Cancer Society. http://www.cancer.org/acs/groups/content/@epidemiologysurveilance/documents/document/acspc-031941.pdf. Accessed May 15, 2013.

2. Colorectal Cancer Facts & Figures 2011–2013 [Internet]. American Cancer Society. http://www.cancer.org/acs/groups/content/@epidemiologysurveilance/documents/document/acspc-028323.pdf. Accessed May 15, 2013.

3. Yabroff KR, Lund J, Kepka D, et al. Economic burden of cancer in the United States: estimates, projections, and future research. Cancer Epidemiol Biomarkers Prev. 2011 Oct;20(10):2006-14. PMID: 21980008.

4. Mariotto AB, Yabroff KR, Shao Y, et al. Projections of the cost of cancer care in the United States: 2010–2020. J Natl Cancer Inst. 2011 Jan 19;103(2):117-28. PMID: 21228314.

5. Cheng X, Chen VW, Steele B, et al. Subsite-specific incidence rate and stage of disease in colorectal cancer by race, gender, and age group in the United States, 1992–1997. Cancer 2001 Nov 15;92(10):2547-54. PMID: 11745188.

6. Cheng L, Eng C, Nieman LZ, et al. Trends in colorectal cancer incidence by anatomic site and disease stage in the United States from 1976 to 2005. Am J Clin Oncol. 2011 Dec;34(6):573-80. PMID: 21217399.

7. Screening for Colorectal Cancer [Internet]. U.S. Preventive Task Force. http://www.uspreventiveservicestaskforce.org/uspstf/uspscolo.htm. Accessed January 23, 2013.

8. National Comprehensive Cancer Network. NCCN Guidelines in oncology: colon cancer V3.2013 [slide set]. Fort Washington (PA): National Comprehensive Cancer Network (NCCN); 2012 Nov 26. 120 p. Also available: http://www.nccn.org/professionals/physician_gls/pdf/colon.pdf.

9. Fazel R, Krumholz HM, Wang Y, et al. Exposure to low-dose ionizing radiation from medical imaging procedures. N Engl J Med. 2009 Aug 27;361(9):849-57. PMID: 19710483.

10. Dewhurst C, Rosen MP, Blake MA, et al. ACR Appropriateness Criteria pretreatment staging of colorectal cancer. J Am Coll Radiol. 2012 Nov;9(11):775-81. PMID: 23122343.

11. Blake MA, McDermott S, Rosen MP, et al; Expert Panel on Gastrointestinal Imaging. American College of Radiology Appropriateness Criteria®: Suspected Liver Metastases [Online Publication]. American College of Radiology. http://gm.acr.org/SecondaryMainMenuCategories/quality_safety/app_criteria/pdf/ExpertPanelonGastrointestinalImaging/SuspectedLiverMetastasesDoc14.aspx. Accessed May 15, 2013.

12. Owens DK, Lohr KN, Atkins D, et al. Grading the strength of a body of evidence when comparing medical interventions—Agency for Healthcare Research and Quality and the Effective Health -Care Program. J Clin Epidemiol. 2010 May;63(5):513-23. PMID: 19595577.

13. Chalmers I, Adams M, Dickersin K, et al. A cohort study of summary reports of controlled trials. JAMA. 1990 Mar 9;263(10):1401-5. PMID: 2304219.

14. Neinstein LS. A review of Society for Adolescent Medicine abstracts and *Journal of Adolescent Health Care* articles. J Adolesc Health Care. 1987 Mar;8(2):198-203. PMID: 3818406.

15. Dundar Y, Dodd S, Williamson P, et al. Case study of the comparison of data from conference abstracts and full-text articles in health technology assessment of rapidly evolving technologies: does it make a difference? Int J Technol Assess Health Care. 2006 Summer;22(3):288-94. PMID: 16984055

16. De Bellefeuille C, Morrison CA, Tannock IF. The fate of abstracts submitted to a cancer meeting: factors which influence presentation and subsequent publication. Ann Oncol. 1992 Mar;3(3):187-91. PMID: 1586615.

17. Scherer RW, Langenberg P, von Elm E. Full publication of results initially presented in abstracts. Cochrane Database Syst Rev. 2007 Apr 18:(2):MR000005. PMID: 17443628.

18. Yentis SM, Campbell FA, Lerman J. Publication of abstracts presented at anaesthesia meetings. Can J Anaesth. 1993 Jul;40(7):632-4. PMID: 8403137.

19. Marx WF, Cloft HJ, Do HM, et al. The fate of neuroradiologic abstracts presented at national meetings in 1993: rate of subsequent publication in peer-reviewed, indexed journals. AJNR Am J Neuroradiol. 1999 Jun-Jul;20(6):1173-7. PMID: 10445467.

20. Moher D, Pham B, Klassen TP, et al. What contributions do languages other than English make on the results of meta-analyses? J Clin Epidemiol. 2000 Sep;53(9):964-72. PMID: 11004423.

21. Juni P, Holenstein F, Sterne J, et al. Direction and impact of language bias in meta-analyses of controlled trials: empirical study. Int J Epidemiol. 2002 Feb;31(1):115-23. PMID: 11914306.

22. Li FY, Lai MD. Colorectal cancer, one entity or three? J Zhejiang Univ Sci B. 2009 Mar;10(3):219-29. PMID: 19283877.

23. Gowdra Halappa V, Corona Villalobos CP, Bonekamp S, et al. Rectal imaging: Part 1, high-resolution MRI of carcinoma of the rectum at 3 T. AJR Am J Roentgenol. 2012 Jul;199(1):W35-42. PMID: 22733930.

24. Whiting P, Rutjes AW, Reitsma JB, et al. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol. 2003 Nov 10;3(1):25. PMID: 14606960.

25. Lijmer JG, Mol BW, Heisterkamp S, et al. Empirical evidence of design-related bias in studies of diagnostic tests. JAMA. 1999 Sep 15;282(11):1061-6. PMID: 10493205.

26. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011 Oct 18;155(8):529-36. PMID: 22007046.

27. Treadwell JR, Tregear SJ, Reston JT, et al. A system for rating the stability and strength of medical evidence. BMC Med Res Methodol. 2006 Oct 19;6:52. PMID: 17052350.

28. Viswanathan M, Ansari MT, Berkman ND, et al. Assessing the risk of bias of individual studies in systematic reviews of health care interventions. In: Methods Guide for Effectiveness and Comparative Effectiveness Reviews. AHRQ Publication No. 10(12)-EHC063-EF. Rockville, MD: Agency for Healthcare Research and Quality; 2012:69-97.

29. White CM, Ip S, McPheeters M, et al. Using existing systematic reviews to replace de novo processes in conducting comparative effectiveness reviews. In: Methods Guide for Effectiveness and Comparative Effectiveness Reviews. AHRQ Publication No. 10(12)-EHC063-EF. Rockville, MD: Agency for Healthcare Research and Quality; 2012:163-78.

30. Harbord RM, Deeks JJ, Egger M, et al. A unification of models for meta-analysis of diagnostic accuracy studies. Biostatistics. 2007 Apr;8(2):239-51. PMID: 16698768.

31. STATA®/MP Data Analysis and Statistical Software: Single User STATA for Windows. College Station, TX: StataCorp; 1996-2013. http://www.stata.com/statamp/. Accessed May 15, 2013. 32. Zamora J, Abraira V, Muriel A, et al. Meta-DiSc: a software for meta-analysis of test accuracy data. BMC Med Res Methodol. 2006 Jul 12;6:31. PMID: 16836745.

33. Chang SM, Matchar DB, eds. Methods Guide for Medical Test Reviews. AHRQ Publication No. 12-EHC017. Rockville, MD: Agency for Healthcare Research and Quality; June 2012.

34. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med. 2002 Jun 15;21(11):1539-58. PMID: 12111919.

35. Higgins JP, Thompson SG, Deeks JJ, et al. Measuring inconsistency in meta-analyses. BMJ. 2003 Sep 6;327(7414):557-60. PMID: 12958120.

36. Shea BJ, Grimshaw JM, Wells GA, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Med Res Methodol. 2007 Feb 15;7:10. PMID: 17302989.

VI. Definition of Terms

Staging. The process of categorizing tumors as per standardized classification systems based on the size, extent, and spread of disease. Staging determines subsequent treatment and is also a prognostic factor for patients with colorectal cancer.

Restaging. The process of using imaging to assess the extent and spread of disease before surgery (but after initial treatment) in patients with colorectal cancer who receive chemotherapy and/or radiotherapy before surgery.

Stage reclassification. The change in stage of the cancer after imaging, compared with the stage determined based on information available before imaging.

VII. Summary of Protocol Amendments

In the event of protocol amendments, the date of each amendment will be accompanied by a description of the change and the rationale. Changes made to the protocol will not be incorporated throughout the various sections of the protocol. Instead, protocol amendments will only be noted in section VII of the protocol, and the date of the amendment noted at the top of the protocol.

This is the first draft of the protocol, and no amendments have yet been made.

VIII. Review of Key Questions

Key Questions were reviewed and refined as needed by the EPC with input from Key Informants and the TEP to ensure that the questions are specific and explicit about what information is being reviewed. Additionally, the Key Questions were posted for public comment and finalized by the EPC after review of the comments.

IX. Key Informants

Key Informants are the end-users of research, including patients and caregivers, practicing clinicians, relevant professional and consumer organizations, purchasers of health care, and others with experience in making health care decisions. Within the EPC program, the Key Informant role is to provide input into identifying the Key Questions for research that will inform healthcare decisions. The EPC solicits input from Key Informants when developing questions for systematic review or when identifying high-priority research gaps and needed new research. Key Informants are not involved in analyzing the evidence or writing the report and have not reviewed the report, except as given the opportunity to do so through the peer or public review mechanism.

Key Informants must disclose any financial conflicts of interest of more than $10,000 and any other relevant business or professional conflicts of interest. Because of their role as end-users, individuals are invited to serve as Key Informants and those who have potential conflicts

may be retained. The Task Order Officer (TOO) and the EPC work to balance, manage, or mitigate any potential conflicts of interest identified.

X. Technical Experts

Technical Experts comprise a multidisciplinary group of clinical, content, and methodological experts who provide input in defining populations, interventions, comparisons, or outcomes as well as identifying particular studies or databases to search. They are selected to provide broad expertise and perspectives specific to the topic under development. Divergent and conflicted opinions are common and perceived as healthy scientific discourse that results in a thoughtful, relevant systematic review. Therefore study questions, design, and/or methodological approaches do not necessarily represent the views of individual technical and content experts. Technical Experts provide information to the EPC to identify literature search strategies and recommend approaches to specific issues as requested by the EPC. Technical Experts neither do analysis of any kind nor contribute to writing the report and have not reviewed it, except as given the opportunity to do so through the peer or public review mechanism.

Technical Experts must disclose any financial conflicts of interest of more than $10,000 and any other relevant business or professional conflicts of interest. Individuals are invited to serve as Technical Experts because of their unique clinical or content expertise and those who have potential conflicts may be retained. The TOO and the EPC work to balance, manage, or mitigate any potential conflicts of interest identified.

XI. Peer Reviewers

Peer Reviewers are invited to provide written comments on the draft report based on their clinical, content, or methodological expertise. Peer review comments on the preliminary draft of the report are considered by the EPC in preparing the final draft of the report. Peer reviewers do not participate in writing or editing the final report or other products. The synthesis of the scientific literature presented in the final report does not necessarily represent the views of individual reviewers. The dispositions of the peer review comments are documented and will, for CERs and Technical Briefs, be published 3 months after the publication of the Evidence Report.

Potential reviewers must disclose any financial conflicts of interest of more than $10,000 and any other relevant business or professional conflicts of interest. Invited Peer Reviewers may not have any financial conflict of interest of more than $10,000. Peer Reviewers who disclose potential business or professional conflicts of interest may submit comments on draft reports through the public comment mechanism.

XII. EPC Team Disclosures

EPC core team members must disclose any financial conflicts of interest of more than $1,000 and any other relevant business or professional conflicts of interest. Related financial conflicts of interest that cumulatively total more than $1,000 will usually disqualify EPC core team investigators.

XIII. Role of the Funder

This project was funded under Contract No. HHSA 290-2012-00011-I from AHRQ, U.S. Department of Health and Human Services. The TOO reviewed contract deliverables for adherence to contract requirements and quality. The authors of this report are responsible for its content. Statements in the report should not be construed as endorsement by the AHRQ or the U.S. Department of Health and Human Services.

Appendix A. Colorectal Cancer Staging Systems

Appendix Table 1: Tumor-Node-Metastasis (TNM) definitions for colorectal cancer

T	N	M
Tx: No description of the tumor's extent is possible because of incomplete information. **Tis:** The cancer is in the earliest stage (in situ). It involves only the mucosa. It has not grown beyond the muscularis mucosa (inner muscle layer). **T1:** The cancer has grown through the muscularis mucosa and extends into the submucosa. **T2:** The cancer has grown through the submucosa and extends into the muscularis propria (thick outer muscle layer). **T3:** The cancer has grown through the muscularis propria and into the outermost layers of the colon or rectum but not through them. It has not reached any nearby organs or tissues. **T4a:** The cancer has grown through the serosa (also known as the visceral peritoneum), the outermost lining of the intestines. **T4b:** The cancer has grown through the wall of the colon or rectum and is attached to or invades into nearby tissues or organs.	**Nx:** No description of lymph node involvement is possible because of incomplete information. **N0:** No cancer in nearby lymph nodes. **N1:** Cancer cells are found in or near 1 to 3 nearby lymph nodes **N1a:** Cancer cells are found in 1 nearby lymph node. **N1b:** Cancer cells are found in 2 to 3 nearby lymph nodes. **N1c:** Small deposits of cancer cells are found in areas of fat near lymph nodes, but not in the lymph nodes themselves. **N2:** Cancer cells are found in 4 or more nearby lymph nodes **N2a:** Cancer cells are found in 4 to 6 nearby lymph nodes. **N2b:** Cancer cells are found in 7 or more nearby lymph nodes.	**M0:** No distant spread is seen. **M1a:** The cancer has spread to 1 distant organ or set of distant lymph nodes. **M1b:** The cancer has spread to more than 1 distant organ or set of distant lymph nodes, or it has spread to distant parts of the peritoneum (the lining of the abdominal cavity).

T: Categories of colorectal cancer describe the extent of spread through the layers that form the wall of the colon and rectum

N: Categories indicate whether or not the cancer has spread to nearby lymph nodes and, if so, how many lymph nodes are involved

M: Categories indicate whether or not the cancer has spread (metastasized) to distant organs, such as the liver, lungs, or distant lymph nodes.

Appendix Table 2. Dukes System

A	Tumor confined to the intestinal wall
B	Tumor invading through the intestinal wall
C1	With lymph node involvement but not apical node
C2	With lymph node involvement including apical node
D	Distant metastasis

Appendix Table 3. Modified Astler-Coller System

A	Tumor limited to mucosa
B1	Tumor invading into muscularis
B2	Tumor invading into serosa
B3	Tumor invading into adjacent organs
C1, C2, C3	Relevant B category but with lymph node involvement
D	Distant metastasis

Appendix Table 4. Taxonomic and prognostic groups based on the AJCC, Dukes and Modified Astler-Coller staging systems

Stage	T	N	M	Dukes	MAC
0	Tis	N0	M0	—	—
I	T1	N0	M0	A	A
	T2	N0	M0	A	B1
IIA	T3	N0	M0	B	B2
IIB	T4a	N0	M0	B	B2
IIC	T4b	N0	M0	B	B3
IIIA	T1-T2	N1/N1c	M0	C	C1
	T1	N2a	M0	C	C1
IIIB	T3-T4a	N1/N1c	M0	C	C2
	T2-T3	N2a	M0	C	C1/C2
	T1-T2	N2b	M0	C	C1
IIIC	T4a	N2a	M0	C	C2
	T3-T4a	N2b	M0	C	C2
	T4b	N1-N2	M0	C	C3
IVA	Any T	Any N	M1a	D	D
IVB	Any T	Any N	M1b	D	D

AJCC: American Joint Committee on Cancer
MAC: Modified Astler-Coller system

Appendix B. Sample Search Strategy for EMBASE® and MEDLINE® (OVID)

Appendix Table 5. Sample Search Strategy for EMBASE® and MEDLINE® (OVID)

Set Number	Concept	Search Statement
1	Colorectal Cancer	exp Colorectal Neoplasms/ or exp colon cancer/ or exp colon tumor/ or exp rectum cancer/ or exp rectum tumor/ or ((Colon$ or colorectal or rect$) adj2 (cancer$ or tumo$ or neoplas$ or carcinoma$ or adenocarcinoma$)).ti,ab.
2	Staging	neoplasm staging/ or cancer staging/ or (stag$ or restag$ or re-stag$).ti,ab.
3	Imaging Controlled Vocabulary	exp Diagnostic Imaging/ or exp Tomography, Emission-Computed/ or exp Tomography, X-Ray Computed/ or exp Magnetic Resonance Imaging/ or exp Ultrasonography/ or Radiography, Thoracic/ or exp computer assisted tomography/ or positron emission tomography/ or multidetector computed tomography/ or exp nuclear magnetic resonance imaging/ or Thorax radiography/ or exp echography/ or computer assisted emission tomography/ or Endoscopy, Gastrointestinal/ or gastrointestinal endoscopy/ or ("computed tomography" or "computerized tomography" or "multidetector computerized tomography" or "magnetic resonance imaging" or "positron emission tomography" or (CT or PET or MRI or TRUS or TUS or ERUS or EUS or MD-CT or x-ray) or ((endorectal or endoscop$ or transrectal or transabdominal) and ultrasound) or imag$).mp
4	Combine	1 and 2 and 3
5	English	limit 4 to english language
6	Human	limit 5 to human
7	1980–2013	limit 6 to yr="1980 - 2013"
8	Humans	limit 7 to humans
9	Publication Types	8 not (letter/ or editorial/ or news/ or comment/ or case report.mp. or case reports/ or note/ or conference paper/ or (letter or editorial or news or comment or case reports or conference abstract$).pt.)
10	Publication Types	8 and case series
11	Combine	9 or 10
12	Dedupe	remove duplicates from 11

Appendix C. Instruments for Assessing Methodological Risk of Bias of Individual Studies and Systematic Reviews

Modified QUADAS Instrument[24] for diagnostic test performance studies

1. Did the study enroll all, consecutive, or a random sample of patients?
2. Were more than 85 percent of the approached/eligible patients enrolled?
3. Were the patient inclusion/exclusion criteria applied consistently to all patients?
4. Was the study affected by obvious spectrum bias?
5. Did the study account for inter-reader differences?
6. Were readers of the diagnostic test of interest blinded to the results of the reference standard?
7. Were readers of the reference standard blinded to the results of the diagnostic test of interest?
8. Were readers of the diagnostic test of interest blinded to all other clinical information?
9. Were readers of the reference standard blinded to all other clinical information?
10. Were patients assessed by a reference standard regardless of the test's results?
11. Were all patients assessed by the same reference standard regardless of the test's results?
12. If the study reported data for a single diagnostic threshold, was the threshold chosen *a priori*?
13. Were the study results unaffected by intervening treatments or disease progression/regression?
14. Were at least 85 percent of the enrolled patients accounted for?
15. Was the funding for the study derived from a source that does not have a financial interest in its results?

ECRI Instrument for controlled/comparative studies

1. Were patients randomly assigned to the study's groups?
2. Did the study use appropriate randomization methods?
3. Was there concealment of group allocation?
4. For nonrandomized trials, did the study employ any other methods to enhance group comparability?
5. Was the process of assigning patients to groups made independently from physician and patient preference?
6. Did the patients in different study groups have similar levels of performance on the outcome of interest at the time they were assigned to groups?
7. Were the study groups comparable for all other important factors at the time they were assigned to groups?
8. Did the study enroll all suitable patients or consecutive suitable patients?
9. Was the comparison of interest prospectively planned?

10. If the patients received ancillary treatment(s), was there a ≤ 5-percent difference between groups in the proportion of patients receiving each specific ancillary treatment?
11. Were the two groups treated concurrently?
12. Was compliance with treatment ≥ 85 percent in both of the study's groups?
13. Were patients blinded to the treatment they received?
14. Was the healthcare provider blinded to the groups to which the patients were assigned?
15. Were those who assessed the patient's outcomes blinded to the group to which the patients were assigned?
16. Was the integrity of blinding of patients, physicians, or outcome assessors tested and found to be preserved?
17. Was the outcome measure of interest objective and was it objectively measured?
18. Was a standard instrument used to measure the outcome?
19. Was there a ≤ 15-percent difference in the length of followup for the two groups?
20. Did ≥ 85 percent of the patients complete the study?
21. Was there a ≤ 15-percent difference in completion rates in the study's groups?
22. Was the funding for this study derived from a source that would not benefit financially from results in a particular direction?

Modified AMSTAR Instrument[29,36] for systematic reviews

1. Was an *a priori* design or protocol provided?
2. Was a comprehensive search strategy performed?

2a. Was this strategy appropriate to address the relevant Key Question of the CER?

3. Was a list of included and excluded studies provided?
4. Was the application of inclusion/exclusion criteria unbiased?

4a. Are the inclusion/exclusion criteria appropriate to address the relevant Key Question of the CER?

5. Was there duplicate study selection and data extraction?
6. Were the characteristics of the included studies provided?
7. Was the individual study quality assessed?

7a. Was the method of study quality assessment consistent with that recommended by the *Methods Guide for Effectiveness and Comparative Effectiveness Reviews*?

7b. Was the scientific quality of the individual studies used appropriately in formulating conclusions?

8. Were the methods used to combine the findings of studies appropriate?
9. Was the likelihood of publication bias assessed?
10. Have the authors disclosed conflicts of interest?